

# SAR Tomography via Nonlinear Blind Scatterer Separation

Yuanyuan Wang, *Member, IEEE*, and Xiao Xiang Zhu, *Senior Member, IEEE*

**Abstract**—Layover separation has been fundamental to many synthetic aperture radar applications, such as building reconstruction and biomass estimation. Retrieving the scattering profile along the mixed dimension (elevation) is typically solved by inversion of the SAR imaging model, a process known as SAR tomography. This paper proposes a nonlinear blind scatterer separation method to retrieve the phase centers of the layovered scatterers, avoiding the computationally expensive tomographic inversion. We demonstrate that conventional linear separation methods, e.g., principle component analysis (PCA), can only partially separate the scatterers under good conditions. These methods produce systematic phase bias in the retrieved scatterers due to the nonorthogonality of the scatterers steering vectors, especially when the intensities of the sources are similar or the number of images is low. The proposed method artificially increases the dimensionality of the data using kernel PCA, hence mitigating the aforementioned limitations. In the processing, the proposed method sequentially deflates the covariance matrix using the estimate of the brightest scatterer from kernel PCA. Simulations demonstrate the superior performance of the proposed method over conventional PCA-based methods in various respects. Experiments using TerraSAR-X data show an improvement in height reconstruction accuracy by a factor of one to three, depending on the used number of looks.

**Index Terms**—blind source separation; kernel PCA; multibaseline InSAR; nonlinear kernel; SAR tomography.

## I. INTRODUCTION

**S**YNTHETIC aperture radar (SAR) interferometry is by far the most popular method for obtaining global digital elevation models, as well as for assessing long-term millimeter-level deformation over large areas of Earth's surface. However, its side-looking imaging geometry causes inevitable layover in SAR images, especially in mountainous and urban areas. As an example, Fig. 1 (left) shows the layover phenomenon in a dense urban area. The buildings are layovered with the ground in front of them, appearing as if collapsed towards the sensor. As a result, the backscattering from the building facade and the ground is integrated into a single pixel during SAR

image formation. Separating the contributions from different scatterers within one resolution cell has been the fundamental to many applications, such as urban 3-D reconstruction, and deformation monitoring in mountainous areas. It is typically solved by explicit inversion of the SAR imaging model in order to retrieve the scattering profile along the mixed dimension that is *elevation*. This process is known as SAR tomography (TomoSAR) or differential TomoSAR (D-TomoSAR), when the temporal motion of individual scatterers is also considered.

To further describe the layover effect, Fig. 1 (right) illustrates the SAR imaging model at a fixed azimuth position, where  $\gamma(s)$  is the reflectivity profile along the elevation  $s$ . TomoSAR builds a synthetic aperture along  $s$  by utilizing multiple observations at different baselines. The multiple observations are usually acquired in a repeat-pass manner for spaceborne SAR sensors. The continuous form of the SAR imaging model can be expressed as a Fourier transform at discrete frequencies, as follows.

$$g_n = \int_s \gamma(s) \exp\left(-j \frac{4\pi B_n}{\lambda r} s\right) ds, \quad (1)$$

where  $g_n$  is the complex-valued observation acquired at baseline  $B_n$ ,  $\lambda$  and  $r$  are the radar wavelength and range to the object, respectively. Equation (1) can be discretized as

$$\mathbf{g} = \mathbf{R}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad (2)$$

where  $\mathbf{g} \in \mathbb{C}^N$  is the observation vector with  $N$  elements,  $\mathbf{R} \in \mathbb{C}^{N \times L}$  is the so-called *steering matrix*, a discrete Fourier transform depending on the baselines and the  $L$  discrete elevation values,  $\boldsymbol{\gamma} \in \mathbb{C}^L$  is the discretized reflectivity profile, and  $\boldsymbol{\varepsilon} \in \mathbb{C}^N$  is the noise vector usually assumed to be independent and identically distributed (i.i.d.) complex circular Gaussian random variables. Due to its layover separation capability, TomoSAR has become the most competent multibaseline SAR interferometry (InSAR) techniques for various 3D reconstruction tasks. The technique has undergone extensive development since the era of very high-resolution spaceborne SAR sensors, e.g., TerraSAR-X. Some examples include improving the estimator by introducing regularization [1], [2] using singular value decomposition, including nonlinear deformation parameters [3], [4], [5], improving the detection of scatterers [6], [7], employing sparse reconstruction techniques to achieve super-resolution in the elevation reconstruction [8], [9], as well as fusing SAR imaging geodesy [10] and TomoSAR inversion to obtain absolute geodetic TomoSAR [11] point clouds. Comprehensive reviews of TomoSAR algorithms can be found in [9], [12], [13].

This research was jointly supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. [ERC-2016-StG-714087], Acronym: *So2Sat*), by the Helmholtz Association through the Framework of Helmholtz Artificial Intelligence Cooperation Unit (HAICU) - Local Unit "Munich Unit @Aeronautics, Space and Transport (MASTr)" and Helmholtz Excellent Professorship "Data Science in Earth Observation - Big Data Fusion for Urban Research" and by the German Federal Ministry of Education and Research (BMBF) in the framework of the international future AI lab "AI4EO - Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond" (*Corresponding Author: Xiao Xiang Zhu*)

Y. Wang and X. Zhu are with the Signal Processing in Earth Observation, Technical University of Munich, Munich, Germany, as well as with the Remote Sensing Technology Institute, German Aerospace Center. (email: yuanyuan.wang@dlr.de; xiaoxiang.zhu@dlr.de)

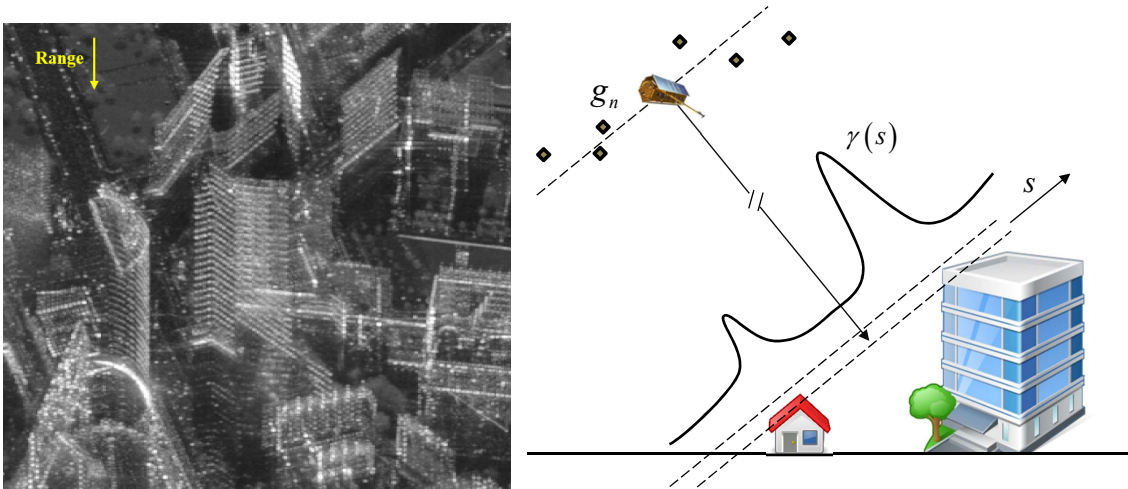


Fig. 1: Left: a SAR intensity image showing the severe layover effect in a dense urban area (Berlin, Germany). The building “collapsed” towards the range direction. Right: the SAR imaging model at a fixed azimuth position. TomoSAR retrieves the reflectivity profile  $\gamma(s)$  by building a synthetic aperture along the direction  $s$  using multiple acquisitions (indicated by the black diamonds) at different baselines.

High precision SAR tomographic reconstruction, and those requiring superresolution is computationally expensive than those traditional methods like persistent scatterer interferometry (PSI). However, future SAR data will eventually converge to high resolution and wide coverage. For example, the German X-band high resolution wide swath SAR sensor is planned to launch in 2022 [14]. At 0.25m resolution, it can cover  $30 \times 30 \text{ km}^2$ , which is tens of times larger than the coverage of the current staring spotlight mode of TerraSAR-X. Therefore, there is an emerging demand for developing computationally economical TomoSAR algorithms. This paper seeks to answer this question by summarizing the state of the art, enumerating the challenges, and proposing a new algorithm as well as directions for future development.

#### A. Related work

The higher computational cost of D-TomoSAR relative to PSI is due to its consideration of multiple scatterers instead of a single one, as in PSI. In D-TomoSAR, the parameters of all scatterers, which are the elevations and deformation parameters, must be searched simultaneously, exponentially increasing the solution space. In a rough approximation, the computational cost of D-TomoSAR is in the order of  $O(L^K)$ , assuming  $O(L)$  is the computational cost of a single-scatterer periodogram in PSI,  $L$  is the discretization level of the parameter space, and  $K$  is the number of scatterers.

Reducing the computational cost of multi-dimensional optimization is often addressed by decomposing the optimization into several sub-problems whose optimization can be performed independently. This has been reflected in TomoSAR methods employing principle component analysis (PCA), such as the CAESAR algorithm [15]. By separating the contributions from different scatterers, the computational complexity can be theoretically reduced to  $O(KL)$ .

In a more general context, this type of decomposition problem is regarded as *blind source separation* (BSS), which sep-

arates the contribution of individual sources without knowing the mixing matrix. In simple linear mixing case  $\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}$ , the goal of BSS is to retrieve both the mixing matrix  $\mathbf{A}$  and the source  $\mathbf{s}$ , given only the observations  $\mathbf{x}$  subject to unknown measurement noise  $\mathbf{n}$ . As an alternative to model-based approaches, BSS methods have emerged as data-driven approaches in data science, where the unknown dynamics of the data are often hard to characterize. The most well-known BSS algorithms are inarguably those exploiting second order statistics of the data using PCA [16], [17], and those exploiting higher order statistics using independent component analysis (ICA) [18], [19]. Extension has been created using kernel PCA (KPCA) [20] to tackle nonlinear mixing models [21], [22], [23]. A great deal of the attention has also been devoted to the joint diagonalization of a set of matrices [19], [24], [25], because the underlying key features of the mixed sources, e.g., statistical independence, can be expressed in terms of diagonal matrices [26]. For example, PCA is essentially a diagonalization of the data covariance matrix. The list of literature on BSS is extensive, since the application of BSS covers vast research fields, including hyperspectral imaging, medical imaging, radar, electroencephalogram, audio processing, chemometrics, and so on.

Only a handful of studies has addressed BSS in InSAR. Most of them deal with the problem of polarimetric target decomposition using PCA and ICA [27], [28]. In the context of multibaseline InSAR, [15] proposes using PCA to decompose layovered scatterers. However, [29] argues that it is difficult to assign a physical interpretation to the eigenvectors of a data covariance matrix unless only a single dominant scatterer is present. Nevertheless, a reasonable result was demonstrated in [15]. We discovered that PCA can only be applied in certain conditions, and proposed to employ KPCA in order to mitigate the errors [30].

### B. Contribution of this paper

The BSS is different in SAR tomography than in conventional BSS applications, because the signal of our interest is often only the phase, instead of the whole complex-value. There has not been a systematic study of blindly separating complex multibaseline InSAR signals. The contribution of this paper is as follows.

- 1) It provides a comprehensive review of the state of the art and systematically demonstrates the limitations of conventional methods in scatterers separation, mainly the low orthogonality between the scatterers.
- 2) It proposes a nonlinear method to mitigate the phase bias caused by the limitations of the state-of-the-art algorithms. The proposed method employs KPCA to iteratively retrieve the direction of the currently most dominant scatterer. The contribution of the dominant scatterer in each iteration is deflated from the covariance matrix after estimating the intensity using the Rayleigh quotient. The algorithm is fully nonparametric.
- 3) We also designed a robust workflow for real data processing.

### C. Notations

This paper make use of a bold capital letter to denote a matrix, bold lowercase letter for a vector, and italic letters (both upper and lowercase) for a scalar. Frequently appearing quantities and notations in the paper are listed in Table I.

TABLE I: Symbols and notations

$N \in \mathbb{N}$	Number of observations (images in our case)
$M \in \mathbb{N}$	Number of samples
$L \in \mathbb{N}$	Discretization level in the elevation direction
$K \in \mathbb{N}$	Number of scatterers
$\mathbf{g} \in \mathbb{C}^N$	Multibaseline InSAR observations
$\boldsymbol{\gamma} \in \mathbb{C}^L$	Reflectivity profile along elevation
$\mathbf{G} \in \mathbb{C}^{N \times M}$	$M$ samples of multibaseline InSAR observations
$\mathbf{R} \in \mathbb{C}^{N \times K}$	Forward model matrix (aka. Steering matrix), for $K$ scatterers
$\mathbf{C} \in \mathbb{C}^{N \times N}$	Complex-valued covariance matrix
$\boldsymbol{\Phi} \in \mathbb{C}^{N \times N}$	Diagonal matrix of the modeled phase
$\langle \cdot, \cdot \rangle$	Inner product
$\det(\cdot)$	Matrix determinant
$ \cdot $	Absolute value (element-wise if on vector or matrix)

## II. PROBLEM FORMULATION

### A. Mathematical model

#### Fully coherent model

Assuming the reflectivity profile is coherent over the multiple acquisitions at different baselines or acquisition time, the discrete SAR imaging model can be expressed by equation (2). In an urban area, the profile  $\boldsymbol{\gamma}$  usually consists of a few scatterers. For a  $K$ -scatterer profile, the imaging model can be simplified to

$$\mathbf{g} = [\mathbf{r}_1 \quad \mathbf{r}_2 \quad \cdots \quad \mathbf{r}_K] \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_K \end{bmatrix} + \boldsymbol{\varepsilon}, \quad (3)$$

where  $\mathbf{r}_k$  is the column steering vector corresponding to the  $k$ th scatterer, and  $\gamma_k$  is the complex-valued amplitude of the  $k$ th scatterer.

The covariance matrix of the observations  $\mathbf{g}$  is as follows

$$\begin{aligned} \mathbf{C}_{\mathbf{g}\mathbf{g}} &= \mathbb{E} \{ \mathbf{R} \boldsymbol{\gamma} \boldsymbol{\gamma}^H \mathbf{R}^H + \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^H \} \\ &= \mathbb{R} \mathbb{E} \{ \boldsymbol{\gamma} \boldsymbol{\gamma}^H \} \mathbf{R}^H + \mathbb{E} \{ \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^H \}, \end{aligned} \quad (4)$$

where  $(\cdot)^H$  is the conjugate transpose operator. The profile  $\boldsymbol{\gamma}$  can be assumed to be uncorrelated, which leads  $\mathbb{E} \{ \boldsymbol{\gamma} \boldsymbol{\gamma}^H \}$  to a diagonal matrix with the expected intensity of individual scatterers. The observation covariance matrix can be simplified to

$$\mathbf{C}_{\mathbf{g}\mathbf{g}} = \sum_{k=1}^K \sigma_k^2 \mathbf{r}_k \mathbf{r}_k^H + \sigma_{\varepsilon}^2 \mathbf{I}, \quad (5)$$

where  $\sigma_k^2$  is the expected intensity of the  $k$ th scatterer, and  $\sigma_{\varepsilon}^2 \mathbf{I}$  accounts for the covariance matrix of the noise. Without losing generality, we can assume the steering vectors  $\mathbf{r}_k$  are all normalized.

Such a scattering model can resemble the layover of very coherent DSs (or even PSs). However, the intensities of the scatterers across different spatial samples are not assumed to be deterministic. They are assumed to be Gaussian scatterers so that equation (5) holds. This is quite common in urban areas, where many adjacent scatterers on facades are very coherent over different (temporal and spatial) baselines, yet their intensities are stochastic among the adjacent samples. This model is known in radar jargon as the Swerling II model [31].

### Decorrelating DS model

In a more general case, the decorrelation of the reflectivity profile  $\boldsymbol{\gamma}$  due to geometric or temporal baselines should be considered. For a  $K$ -scatterers case, the imaging model can be formulated as follows:

$$\mathbf{C} = \sum_{k=1}^K \boldsymbol{\Phi}_k \mathbf{C}_k \boldsymbol{\Phi}_k^H + \sigma_{\varepsilon}^2 \mathbf{I}, \quad (6)$$

where  $\boldsymbol{\Phi}_k$  is a diagonal matrix constructed from  $\mathbf{r}_k$ , and  $\mathbf{C}_k$  is the positive real-valued decorrelation covariance matrix of the  $k$ th scatterer. Under this type of model,  $K$  covariance matrices with a total of  $N \times N \times K$  parameters are to be estimated, instead of the only  $K$  intensities in the fully coherent model. The fully coherent model is a special case of the decorrelating model where  $\mathbf{C}_k$  degenerates to all-ones matrices. Solving the full covariance matrices as well as the steering vectors in equation (6) using BSS remains at a challenge. We only found solutions under certain specific conditions, e.g., constant coherence. Therefore, the decorrelating DS model is not considered in this paper.

The following content of this article will be based on the fully coherent model expressed in equation (5). The goal of BSS is to retrieve the steering vectors  $\mathbf{r}_k$  of the scatterers, as well as the intensities  $\gamma_k$ , from the observations  $\mathbf{g}$  without performing TomoSAR, i.e., inverting the SAR imaging model and detecting the scatterer's location. Depending on the applications, permutation, scaling, or a constant common phase

offset of the steering vectors may all lead to valid solutions. For example, all three options are valid for 3D reconstruction, because only the relative position of the two scatterers matters in practice. In the following content, we define that the steering vectors are sorted in a descending order according to their corresponding intensity, and the steering vectors have a unit norm.

### B. PCA in TomoSAR scatterer separation

The common idea of BSS algorithms is to exploit statistical independency of the sources. By assuming statistical independency, one can seek a diagonalization of the covariance matrix of the observation. PCA [32] diagonalizes the data by converting the data into a set of orthogonal basis, known as principle components, whose variances are subsequently maximized. Performing PCA on a data covariance matrix that is semi-positive definite is equivalent to eigenvalue decomposition (EVD). For our BSS problem, we denote the EVD of a covariance matrix of the observations as follows.

$$\mathbf{C}_{gg} = \mathbf{U}\mathbf{D}\mathbf{U}^H, \quad (7)$$

where  $\mathbf{U}$  and  $\mathbf{D}$  are the eigenvectors and the diagonal matrix, respectively.

Although [29] mentions that it is difficult to assign exact meaning to the eigenvectors,  $\mathbf{U}$  can be approximated as the steering vectors of the individual scatterers under a *strong orthogonality condition* among all the scatterers, as will be explained in Section II-C. Such approximation was employed in the CAESAR algorithm [15]. However, if the multibaseline SAR observation has a single dominant scatterer, the eigenvector corresponding to the largest eigenvalue of the data covariance matrix is an unbiased estimate of the steering vector of the scatterer. This has been mentioned in [29], [33]. We show a short proof as follows.

**Proof:** For a single scatterer, its theoretical covariance matrix is of the form  $\mathbf{C}_{gg} = \Phi_1 \mathbf{C}_1 \Phi_1^H$  according to equation (6), where  $\mathbf{C}_1$  is a real-valued decorrelation matrix, and  $\Phi_1$  is the diagonal matrix of the steering vector.

- 1)  $\mathbf{C}_{gg}$  can be expressed as  $\mathbf{C}_{gg} = \Phi_1 \mathbf{C}_1 \Phi_1^H = \Phi_1 (\mathbf{V}\mathbf{D}\mathbf{V}^H) \Phi_1^H$ , where  $\mathbf{V}$  and  $\mathbf{D}$  are the eigenvectors and the eigenvalues matrix of  $\mathbf{C}_1$ , respectively.
- 2) Let  $\mathbf{U} = \Phi_1 \mathbf{V}$ . It can be shown that  $\mathbf{U} = \Phi_1 \mathbf{V}$  is orthogonal, hence  $\mathbf{U}\mathbf{D}\mathbf{U}^H$  is the EVD of  $\mathbf{C}_{gg}$ .
- 3) Since  $\mathbf{C}_1$  is always real symmetric and positive definite,  $\mathbf{V}$  and  $\mathbf{D}$  are both real, according to [34].
- 4) Since  $\mathbf{V}$  is real, the phase of each column of  $\mathbf{U}$  is identical to the diagonal of  $\Phi_1$ .

### C. Limitations of PCA

#### Inseparability of nonorthogonal sources

The limitation of PCA is two-fold. On one hand, it assumes a linear combination of orthogonal basis. Hence, it is unable to fully recover nonorthogonal basis. On the other hand, the directions of the sources are indeterminable if the variance of the individual sources are identical. These can be exemplified

by a 2-D Gaussian mixture shown in Fig. 2. The two subfigures are mixtures of nonorthogonal and orthogonal 2-D Gaussian sources of unit variance, respectively. The solid arrows in the figure are the true directions of the sources, and the dashed arrows are the directions estimated by PCA. The length of the arrows is three times that of the estimated or the true standard deviation. The left subfigure demonstrates that a nonorthogonal mixture of Gaussian sources cannot be separated by linear PCA. Both the direction and the variance of the sources were not correctly estimated. In contrast, the right subfigure shows that orthogonal mixture of Gaussian sources of identical variance can be separated by linear PCA, but subject to a random angle offset.

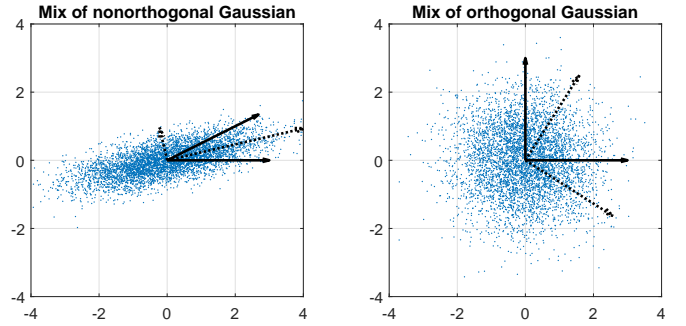


Fig. 2: Mixture of nonorthogonal (left) and orthogonal (right) 2-D Gaussians of identical standard deviation (set to 1). The solid arrows are the true directions of the sources, and the dashed arrows are the directions extracted by PCA. The length of the arrows is three times that of the (estimated) standard deviation. The left subfigure shows that a nonorthogonal mixture of Gaussian sources cannot be separated by linear PCA, while the right subfigure shows that an orthogonal mixture of Gaussian sources with the same variance can also not be unmixed by linear PCA.

Generalizing to our BSS problem in multibaseline InSAR, the performance of PCA is poor when the orthogonality of the two scatterers is low. We discovered the following three common causes of low orthogonality in multibaseline InSAR:

- 1) a low number of images,
- 2) a short distance between the two scatterers, i.e., either close to or shorter than the Rayleigh resolution, and
- 3) similar amplitude among the scatterers, which causes a severe interference.

The degradation of performance is reflected in the phase bias (as well as in the amplitude bias, which is not of our primary interest) of the extracted steering vectors of individual scatterers. An example of this phase bias of the estimated steering vector is shown below.

#### Systematic phase bias

Fig. 3 demonstrates the phase bias of the eigenvectors extracted by PCA from a simulation of a two-scatterer mixture. The number of observations was set to nine. The perpendicular baselines are equally distributed from -200 m to 200 m, with other parameters similar to those of TerraSAR-X. This leads to a Rayleigh resolution of 27m in this simulation.

The two scatterers are set to locate at 40 m and 80 m. No noise was included in the simulation. The two curves in each subplot in Fig. 3 correspond to an amplitude ratio  $\alpha$  of 1.2 or 2 between the two scatterers, respectively. The phase bias appears as a periodic undulation with respect to the perpendicular baseline. Consistent with the analysis shown in Fig. 2 (right), the phase bias increases as the amplitude ratio of the two scatterers approaches one. The bias of the second eigenvector is larger than that of the first. In this example, the maximum phase bias of the eigenvectors for  $\alpha = 2$  is about  $4^\circ$  and  $15^\circ$ . This corresponds to about 0.3m and 1m bias, respectively, in elevation. Such precision is sufficient for certain applications. However, as the amplitude ratio and SNR vary, large systematic phase bias can be expected from conventional PCA-based methods.

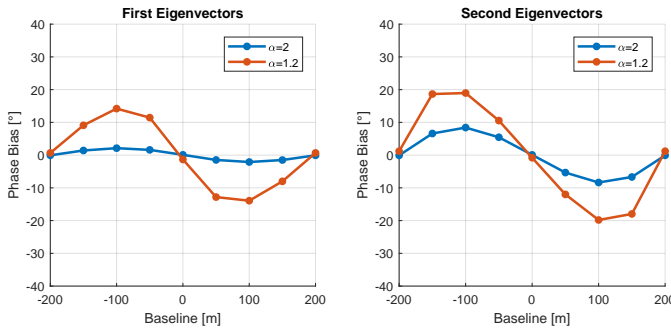


Fig. 3: Phase bias of the first and second eigenvectors with respect to the perpendicular baseline, at different amplitude ratios ( $\alpha = 2, 1.2$ ) of the two scatterers. The larger amplitude ratio between the scatterers gives less phase bias, hence better separability of the scatterers. The bias of the second eigenvector is in general larger than the first one.

### III. NONLINEAR BLIND SCATTERER SEPARATION

Based on the discussion in Section II-C, the performance of PCA degrades as the orthogonality of the two scatterers reduces. As the amplitude ratio of the scatterers in the data cannot be altered, increasing the dimensionality of the data is the fundamental strategy to improve the orthogonality in our proposed algorithm. As it is usually infeasible to increase the number of images, the proposed method artificially increases the dimension of the data by projecting them into a higher dimensional space through nonlinear transformation. The BSS is then performed in the transformed higher dimensional space. The proposed algorithm employs the *kernel trick* [35] to perform the BSS in a Hilbert space without explicitly evaluating in the higher dimensional space. This can be easily realized using KPCA.

The proposed algorithm performs in an iterative manner. At each iteration, it extracts and demodulates the dominant scattering contribution from the data covariance matrix, until no significant scattering is left or the predefined maximum number of scatterers is reached. The flowchart of the proposed algorithm is shown in Fig. 4. The algorithm is explained in detail in the following sections.

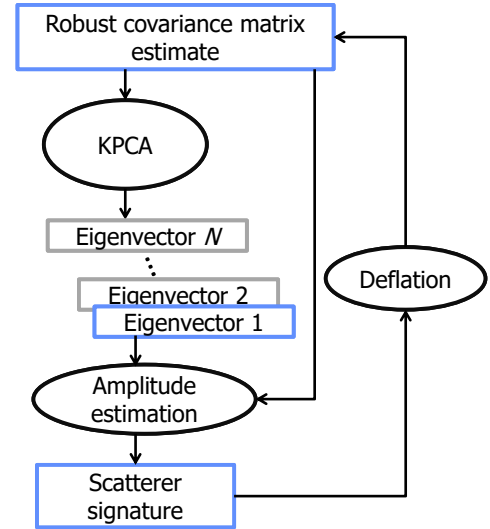


Fig. 4: The flowchart of the proposed nonlinear blind scatterer separation algorithm.

#### A. Dominant scatterer extracting via kernel PCA

We will denote  $\mathbf{G} \in \mathbb{C}^{N \times M}$  as the matrix of  $M$  ergodic samples of the multibaseline observation  $\mathbf{g}$ . The scatterers of each sample are assumed to be stationary random variables, i.e., identical phase centers of the scatterers. In real data, we should imagine pixels with similar layover configurations, e.g., a row of pixels on the same floor of a façade which layovers with flat ground. As the discussion of the sample selection is out of the scope of this paper, we assume the aforementioned condition is met in this article. This allows us to make use of the second order statistics, i.e., the covariance matrix, of the data. Mathematically, the singular vectors of  $\mathbf{G}$  are identical to the eigenvectors of the sample covariance matrix  $\hat{\mathbf{C}} = M^{-1} \mathbf{G} \mathbf{G}^H$ . In the following derivation, we will denote  $\mathbf{C}$  as the input data matrix of BSS, instead of  $\mathbf{G}$ .

PCA performs a linear separation in the original data space, i.e., in  $\mathbf{C}$ . As described above, KPCA is a nonlinear generalization of PCA. The nonlinear separation is achieved through a linear separation on a nonlinearly transformation  $\Psi$  of the data [20], [22], which can be expressed as follows.

$$\Psi: \mathbb{C}^N \rightarrow F, \mathbf{c} \rightarrow \Psi(\mathbf{c}), \quad (8)$$

where  $F$  is the transformed vector space which may have arbitrary dimension, and  $\mathbf{c}$  denotes the columns of  $\mathbf{C}$ . Let us denote the transformed data as  $\Psi_{\mathbf{c}} = [\Psi(\mathbf{c}_1), \Psi(\mathbf{c}_2), \dots, \Psi(\mathbf{c}_N)]$ . The KPCA of  $\mathbf{C}$  basically finds the EVD of the covariance matrix in the transformed space, which is

$$\mathbf{C}_{\Psi\Psi} = \Psi_{\mathbf{c}} \Psi_{\mathbf{c}}^H = \mathbf{U}_{\Psi\Psi} \mathbf{D}_{\Psi\Psi} \mathbf{U}_{\Psi\Psi}^H. \quad (9)$$

As the nonlinear transformation can have infinite dimension, EVD is never explicitly evaluated: they are indirectly evaluated through the kernel trick. It assumes that a Hilbert space of  $F$  can be represented by a certain kernel function of the input data space, i.e.,

$$\kappa(\mathbf{c}_i, \mathbf{c}_j) = \Psi(\mathbf{c}_i)^H \Psi(\mathbf{c}_j), \quad (10)$$



where  $\kappa(\cdot)$  is a kernel function, and  $\mathbf{c}_j$  refers to the  $j$ th sample (column) of  $\mathbf{C}$ . For convenience, we define the kernel matrix  $\mathbf{K} \in \mathbb{C}^{N \times N}$  of the transformed data as

$$\mathbf{K} = \Psi_{\mathbf{c}}^H \Psi_{\mathbf{c}}. \quad (11)$$

Hence, each element of the kernel matrix can be easily found via equation (10). The EVD of the kernel matrix is immediately available as follows.

$$\mathbf{K}\mathbf{V} = \mathbf{V}\mathbf{S}, \quad (12)$$

where  $\mathbf{V}$  is the eigenvectors, and  $\mathbf{S}$  is the matrix of eigenvalues. Multiplying both sides of equation (12) by  $\Psi_{\mathbf{c}}$  gives

$$\Psi_{\mathbf{c}} \Psi_{\mathbf{c}}^H (\Psi_{\mathbf{c}} \mathbf{V}) = (\Psi_{\mathbf{c}} \mathbf{V}) \mathbf{S}, \quad (13)$$

which implies that  $\Psi_{\mathbf{c}} \mathbf{V}$  and  $\mathbf{S}$  are the eigenvectors and eigenvalues of the covariance matrix  $\mathbf{C}_{\Psi\Psi} = \Psi_{\mathbf{c}} \Psi_{\mathbf{c}}^H$ , respectively.

By properly choosing the kernel function,  $\Psi_{\mathbf{c}} \mathbf{V}$  shall represent the space spanned by individual scatterers, or at least one of them. Hence, the data projected onto these eigenvectors shall represent the array manifold, i.e., the steering vectors. By taking and normalizing the first  $K$  eigenvectors of  $\mathbf{C}_{\Psi\Psi}$ , the orthogonal projection basis in the higher dimension can be obtained as follows.

$$\Xi = \Psi_{\mathbf{c}} \mathbf{V}_{1 \sim K} \mathbf{S}_{1 \sim K}^{-1/2}, \quad (14)$$

where  $\mathbf{V}_{1 \sim K}$  and  $\mathbf{S}_{1 \sim K}$  denote the first  $K$  columns of  $\mathbf{V}$  and  $\mathbf{S}$ . Although, the eigenvectors and eigenvalues of  $\mathbf{C}_{\Psi\Psi}$  appear in equation (14), the EVD of  $\mathbf{C}_{\Psi\Psi}$  is never explicitly evaluated. Only the calculation of the kernel matrix is required [22], [36] when projecting the data onto this basis. The projected data can be obtained as follows.

$$\mathbf{Y} = \Psi_{\mathbf{c}}^H \Xi = \mathbf{K} \mathbf{V}_{1 \sim K} \mathbf{S}_{1 \sim K}^{-1/2}. \quad (15)$$

The phase of the first column of  $\mathbf{Y}$  is extracted as the estimate of the steering vector of the first scatterer.

### B. Selection of the kernel

We seek a proper kernel function that allows the energy of different scatterers, or at least the most dominant scatterer, to be well localized in the individual columns of the transformed data  $\mathbf{Y}$ . In order to achieve that, the kernel should be able to transform the elliptically distributed scatterers into a linear coordinate where they can be separated. Kernels that have been extensively discussed in various applications are polynomial and Gaussian kernels, as they are both able to transform a radial basis to a near-linear basis.

### Polynomial kernel

A polynomial kernel is usually defined as follows:

$$\kappa(\mathbf{c}_i, \mathbf{c}_j) = (\mathbf{c}_i^H \mathbf{c}_j + 1)^d, \quad (16)$$

where  $\mathbf{c}_i^H \mathbf{c}_j$  emphasizes the *angular difference* between the data, while the order  $d \in \mathbb{R}^+$  introduces nonlinearity and increases the dimension. For integer polynomial order  $d$ , it computes a dot product in the space spanned by all monomials of degree  $d$  in the input coordinates [22], [35]. For

example, the second order polynomial kernel of a mixture of two sources  $s_1$  and  $s_2$  will be spanned by the monomials  $\{s_1^2 s_2^2, s_1^2 s_2, s_1 s_2^2, s_1^2, s_2^2, s_1 s_2, s_1, s_2\}$ . The dimension under non-integer polynomial orders can be very high. It is obvious that polynomial orders greater than two are not feasible for our BSS problem, since they will introduce artificial scatterers with higher order phase term into the mixture. Such higher order scatterers have effective elevations of multiple times of the original ones, which may cause phase ambiguity.

### Gaussian kernel

A Gaussian kernel is defined in equation (17) below.

$$\kappa(\mathbf{c}_i, \mathbf{c}_j) = \exp\left(-\|\mathbf{c}_i - \mathbf{c}_j\|_2^2 / (2\sigma^2)\right). \quad (17)$$

Unlike the polynomial kernel, a Gaussian kernel emphasizes the Euclidean distance between the data. The standard deviation  $\sigma$  of the Gaussian kernel depends on the *Euclidean distance* between the scatterers. Ideally, it should be smaller than *inter-scatterer* distances while larger than *inner-scatterer* distances. Although, the scatterers are yet unknown, one can still estimate the inner scatterer distance by finding the minimum distance between samples [36]. Following the same idea, we propose an estimator which is expressed in equation (18). The  $\min(\cdot)$  finds the minimum distance among  $\mathbf{c}_j$  to all  $\mathbf{c}_i$ , for  $i = 1, 2, \dots, N, i \neq j$ . The  $\text{mean}(\cdot)$  takes an average over all the minimum distances. The parameter  $\beta$  in the equation is for fine-tuning.

$$\hat{\sigma} = \beta \text{mean}\left(\min_j (\|\mathbf{c}_i - \mathbf{c}_j\|_2)\right). \quad (18)$$

### C. Sequential amplitude estimation and demodulation

We have not found an explicit expression of the secondary eigenvectors out of KPCA with respect to the steering vectors of the scatterers. Therefore, the proposed algorithm can only make use of the first column of  $\mathbf{Y}$  that captures the steering vector of the most dominant scatterer well. The proposed algorithm employs a strategy that sequentially demodulates and estimates the most dominant contribution. However, a challenge arises, where the real variances of the scatterers are lost after the nonlinear transformation in KPCA, i.e., the real intensity of the scatterer is not represented by the eigenvalues of  $\mathbf{Y}$ . Therefore, we estimate the intensity by the Rayleigh quotient [37], i.e.,

$$\hat{\sigma}_1^2 = \frac{1}{N} \frac{\bar{\mathbf{y}}_1^H \hat{\mathbf{C}} \bar{\mathbf{y}}_1}{\bar{\mathbf{y}}_1^H \bar{\mathbf{y}}_1}, \quad (19)$$

where  $\bar{\mathbf{y}}_1$  denotes the first column of  $\mathbf{Y}$  with its amplitude dropped. Please note that the amplitude-dropped vector appears in the equation, instead of the normalized one that usually appears in the literature. Once the intensity of the scatterers is obtained, the covariance matrix can be demodulated as follows:

$$\hat{\mathbf{C}}_{\text{update}} = \hat{\mathbf{C}} - \hat{\sigma}_1^2 \bar{\mathbf{y}}_1 \bar{\mathbf{y}}_1^H. \quad (20)$$

This KPCA plus demodulation process is iteratively performed until no more significant scatterers is left or the predefined maximum number of scatterers is reached.

TABLE II: Summary of the proposed algorithm

---

for each pixel in the image
<b>Sample selection</b> (samples with similar layover configuration)
<b>Estimate sample coherence matrix</b> $\hat{\mathbf{C}}_0 = \frac{1}{M} \mathbf{G} \mathbf{G}^H$
<b>Center</b> $\hat{\mathbf{C}}_0$ yielding $\hat{\mathbf{C}}_0$
<b>Estimate number of scatterer</b> $K$ , e.g., using minimum description length
for $k = \text{scatterer } 1 \text{ to } K$
<b>Perform KPCA</b> on $\hat{\mathbf{C}}_{k-1}$ according to equation (15), obtaining dimension reduced data $\mathbf{Y}$
<b>Obtain estimate of steering vector</b> $\hat{\mathbf{r}}_k$ by taking the first eigenvector (corresponding to the largest eigenvalue) of $\mathbf{Y}$
<b>Estimate scatterer amplitude</b> $\sigma_k$ using Rayleigh quotient (equation (19))
(optional) <b>Re-estimate</b> $\sigma_1$ and $\sigma_2$ alternately when $K$ is only equal to two
<b>Deflate</b> $\hat{\mathbf{C}}_{k-1}$ , i.e., $\hat{\mathbf{C}}_k = \hat{\mathbf{C}}_{k-1} - \hat{\sigma}_k^2 \hat{\mathbf{r}}_k \hat{\mathbf{r}}_k^H$
<b>Center</b> $\hat{\mathbf{C}}_k$ yielding $\hat{\mathbf{C}}_k$
end
<b>Height estimation</b> using standard PSI technique (e.g., periodogram or integer least square) from the estimated steering vectors $[\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2, \dots, \hat{\mathbf{r}}_K]$
end

---

#### D. Algorithm summary

The previous subsection describes the core of the proposed blind TomoSAR algorithm. However, in TomoSAR applications, the processing pipeline shall also perform sample selection, model order selection, and height estimation, besides the main steps of the proposed algorithm. We summarize the full algorithm for a real SAR image processing in Table II. Since the scope of this paper is mainly on mitigate the phase bias in the retrieve the steering vectors, we will not go into details of the other steps. The readers can refer to [38], [39] for sample selection, [40] for robust covariance matrix estimation, [41] for model order selection, and [42] for final parameters estimation.

#### E. Extension to D-TomoSAR

Real data InSAR stacks are often multi-temporal, meaning the deformation shall also be considered. In the case of differential TomoSAR (D-TomoSAR), the reflectivity profile will be  $\gamma(s) \delta(p_1 - p_1(s), \dots, p_M - p_M(s))$  [43], where  $p_i$  is the deformation axis and  $p_i(s)$  is the deformation function along the elevation. The phase of the corresponding steering vectors is the sum of the height frequency as well as the deformation frequency. In discrete form, the dimension of  $\mathbf{R}$  and  $\gamma$  will increase according to the number of the deformation parameters, i.e.,  $\mathbf{R} \in \mathbb{C}^{N \times (L \times P_1 \times P_2 \dots)}$ , and  $\gamma \in \mathbb{C}^{L \times P_1 \times P_2 \dots}$ , where  $P_i$  is the discretization level of the  $i$ th deformation parameter.

Since the height and deformation signal act on a scatterer at a fixed elevation location, the deformation and height signal of a scatterer can be considered as a single source, which can be retrieved using the proposed algorithm without any modification. The only necessary change in the whole processing pipeline is the last step (height estimation) in Table II. It will be a multi-dimensional periodogram, instead of one dimensional. The reader can refer to [30] for the application

of the proposed algorithm on retrieving both the height and periodic deformation induced by thermal dilation.

## IV. EXPERIMENTS

In this section, we evaluate the performance of the proposed method using simulated data, as well as real TerraSAR-X data.

#### A. Simulation

The performance of the proposed algorithm is firstly evaluated via simulations. As we are mainly interested in the phase of the scatterers, we define the angle between the estimated scatterer steering vector and the ground truth as the quality metric. The angle is defined as the arccosine of the inner product of the two vectors, which is shown as follows.

$$b = \cos^{-1} |\hat{\mathbf{r}}^H \mathbf{r}|, \quad (21)$$

where both  $\hat{\mathbf{r}}$  and  $\mathbf{r}$  are assumed to be normalized. Because of the absolute value in equation (21), the domain of the angular bias is  $[0^\circ, 90^\circ]$ . A wild guess of an unknown signal direction will be at  $45^\circ$ . Therefore, an angular bias greater than  $45^\circ$  basically indicates a failed source separation.

The simulation setting was similar to that in Fig. 3. Since most of the TomoSAR literature assumes two dominant scatterers in TerraSAR-X data from urban areas, two-scatterer mixtures were simulated in our experiments. The number of images was set to 9. The baselines are equally distributed from -200m to 200m. The other parameters (e.g., wavelength, range distance, etc.) were set to be similar to those of TerraSAR-X. According to the analysis, the orthogonality and the variance of the signals heavily affects the performance of the BSS algorithms. Therefore, we vary the amplitude ratio, the distance between the two scatterers, and SNR in the simulation. In the first two experiments, i.e. amplitude ratio and scatterer distance, we intend to test the phase bias of PCA and the proposed algorithm at noise free case. Therefore, we used 900 looks to estimate a very accurate covariance matrix. Of course, it is nearly identical by using the theoretical covariance matrix. In all the experiments, we use 1000 times Monte Carlo simulation to estimate the mean and the standard deviation of the angular bias.

#### Performance with respect to amplitude ratio

The first simulation set out to study the systematic bias of BSS algorithms and so no noise was included in the signal. The distance between the two scatterers was set to one Rayleigh resolution, which is 27.3 m in this case. Fig. 5 shows the angular bias of the steering vectors of the two scatterers estimated using PCA and the proposed method. The x-axis shows an increasing amplitude ratio from 1 to 2. As can be seen, very large phase bias appears in the PCA result when the amplitude ratio is low. As the amplitude ratio increases, the dominant direction becomes more prominent, and hence easier for PCA to capture. The same trend appears in the result of both the first and the second scatterers. In summary, PCA will not be a good choice for separating scatterers with similar brightness. However, the performance of PCA will be

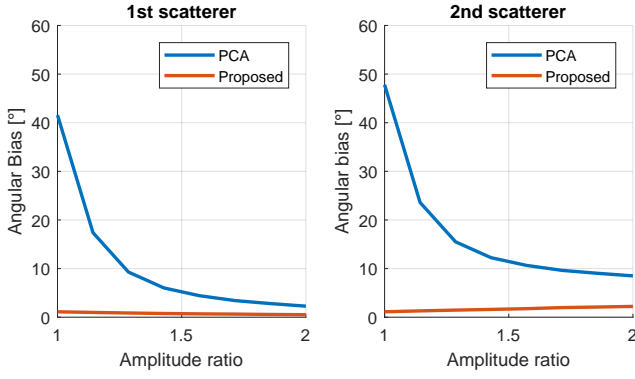


Fig. 5: The phase bias in degree of the estimated steering vectors of the first (brighter) scatterer (left), and the second scatterer (right) with respect to different amplitude ratio between the two scatterers. The elevation difference between the scatterers was set to one Rayleigh resolution (27.3m in the simulation). A Gaussian kernel was employed in the proposed algorithm. A total of 900 looks were used for covariance matrix estimation. No noise was introduced. The figure shows that PCA is sensitive to the intensity ratio of the two scatterers. A comparable intensity between the scatterer will almost result in inseparation of the scatterers, even in a noise-free case.

comparable to the proposed method when the amplitude ratio is larger than 2.

#### Performance with respect to scatterers distance

In this experiment, we varied the distance between the scatterers from 0.3 to 2 Rayleigh resolution units. The amplitude ratio of the scatterers was kept at one. The upper row of Fig. 6 shows the bias of the estimates with respect to increasing distance between the scatterers. For analysis, we also plotted the angle between the true steering vectors of the two scatterers in Fig. 6, which are shown as yellow curves. First, the proposed method greatly outperforms PCA in the whole range of scatterer distance. The angular bias of the first steering vector estimated by the proposed method stays at 2 to 3 degrees, whereas it is 40 degrees for PCA. Second, the performance of both methods stays relatively stable when the distance is larger than 0.8 Rayleigh resolution. Not surprisingly, the performance is clearly affected when it enters the super-resolution regime (i.e., distance shorter than one Rayleigh resolution). Interestingly, the results of PCA have a strong correlation with the original angle of the two scatterers. PCA, as a non-superresolving technique, will detect the first signal at a location between the two scatterers. Hence, the angular bias of the first signal decreases as the distance between the two scatterers decreases. This can be seen in the lower row of Fig. 6, which shows the ratio between the angular bias and the original angle between the two scatterers (steering vectors). In the left plot of the lower row of Fig. 6, the result of PCA stays at 0.5, indicating that PCA always detects the most dominant signal right in the middle of the two scatterers (when they are equally bright).

#### Performance with respect to SNR

The previous experiments show that conventional PCA-based methods have strong systematic bias even under noise-free case. To evaluate the performance of the proposed algorithm under real scenario, we included additional complex circular Gaussian noise in the simulation. As the number of looks and SNR jointly affect the performance, their product was considered. Fig. 7 shows the mean and the standard deviation of the angular bias of the estimates with respect to an increasing  $M \cdot \text{SNR}$ . The solid curves show the performance under a 1:1 amplitude ratio, whereas the dashed curves show the result for the amplitude ratio of 2. First, the result is consistent with Fig. 5 and 6. It is nearly impossible to separate two scatterers using PCA when the amplitude ratio is close to one (red solid curve), regardless of the SNR. The performance of PCA is comparable to the proposed method when the amplitude ratio increases to 2. The SNR mildly affects the performance of the evaluated methods, especially when  $M \cdot \text{SNR}$  is greater than 20dB. In modern high-resolution spaceborne SAR data, a 20 dB  $M \cdot \text{SNR}$  is a rather relaxed condition, for example an SNR of 0dB with 100 looks.

#### B. Performance on real data

We tested our algorithm on six high-resolution TerraSAR-X interferograms acquired in pursuit-monostatic mode. The temporal baselines are in the order of seconds, so that the ground deformation and change in atmospheric phase are negligible. The optical and the SAR image of the test building are shown in Fig. 8. The yellow arrows in the images indicate the range direction. We manually define a iso-height direction parallel to the building floors, in order to guide the sample selection. This direction is shown as the thin red polygon on the amplitude image. We assume that the pixels in the template have similar layover configuration, as well as similar scatterer statistics. This iso-height template slides down one pixel at a time. We estimate a single reflectivity profile from the covariance matrix of the pixels in this template at each given position. In the experiment, we also vary the length of this template to test the performance of the proposed algorithm under different number of looks.

Although this type of iso-height template was shown as an effective sample selection for joint height estimation in TomoSAR [38], the samples within the template cannot be guaranteed to be ergodic. Therefore, a robust covariance matrix estimator was employed in our processing. The sign covariance matrix (SCM) [40] is a non-iterative robust estimator, which is a good balance between robustness and computational cost. The SCM can be estimated as follows:

$$\hat{\mathbf{C}}_{\text{scm}} = \frac{1}{M} \sum_{m=1}^M \|\mathbf{g}_m\|_2^{-2} \mathbf{g}_m \mathbf{g}_m^H, \quad (22)$$

where only the direction of each multivariate sample is considered. The real covariance is lost as a result.

The proposed algorithm was applied to the test building, with the template sliding from the top to the bottom of the building. For each template position, we extract the phase of the two most dominant components as the estimates of steering vectors. For conventional PCA, this will be the first



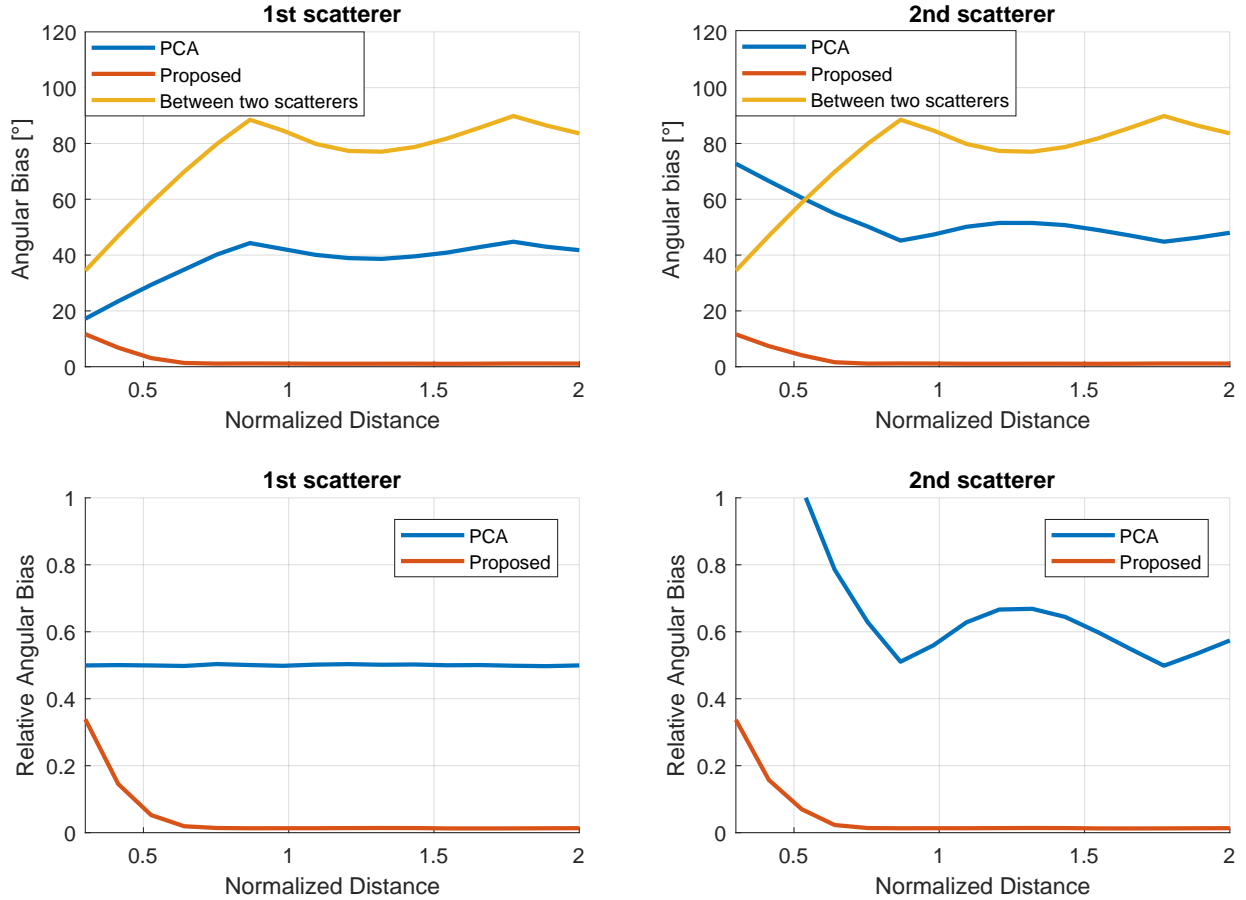


Fig. 6: Upper row: The phase bias (in degree) of the estimated steering vectors of the first (brighter) scatterer (left), and the second scatterer (right) with respect to increasing distance between the two scatterers; lower row: the relative angular bias, which is the ratio of the angular bias and the angle between the true steering vectors of the two scatterers. The amplitude ratio of the scatterers was set to one. A Gaussian kernel was employed in the proposed algorithm. A total of 900 looks were used for covariance matrix estimation. No noise was introduced in the simulation.

two eigenvectors. For each estimate of steering vectors, a periodogram was calculated to estimate the corresponding elevation. In the experiments, we also tested the proposed algorithm at 50 looks and 500 looks, by varying the length of the template.

Figure 9 compares the two-layer elevation retrieved from PCA and the proposed algorithm. The red dots represent the first (brighter) layer, and the blue the second layer. The upper row shows the results using only 50 looks, whereas the lower row is the result using 500 looks. First, it is apparent that the proposed algorithm retrieves a much straighter façade line, whereas small periodic undulations appears on the PCA result. This is likely due to the systematic bias caused by PCA. Second, fewer outliers appear in the results of the proposed method, showing a better robustness against sample nonergodicity. However, the variance of façade from the KPCA result shows slightly higher than that from PCA, indicating the proposed methods requires more number of looks than PCA. By assuming the façade to be a straight line, the accuracy of PCA and the proposed algorithm are both ca. 1.0m using 50 looks. As the number of looks increases, the advantage of the proposed algorithm become more prominent, since the bias of

the estimates becomes more prominent than the variance of the estimates. At 500 looks, the proposed algorithm outperforms PCA by a factor of 1.2 in the accuracy of height estimation. Last but not least, the proposed algorithm was also able to retrieved very well the layover between the top façade and the roof, as well as the layover between the lower façade and the ground (marked by the red ellipse in Fig. 9). The roof-façade layover was also shown in the right of Fig. 8. We can see the minimum distance till which the two layers cannot be separated anymore is roughly 6m, which is below one Rayleigh resolution (11.6m in this data). These findings coincide with the finding in the simulation (Fig. 6) that shows the proposed method has extremely low phase bias, even in the super-resolution regime.

## V. DISCUSSION

### A. Kernel parameter selection

The proposed algorithm requires the selection of a kernel parameter: either the polynomial order  $d$  for a polynomial kernel, or the factor  $\beta$  of the standard deviation in a Gaussian kernel. In order to obtain an operable range of those parameters, we measure the performance at different parameter

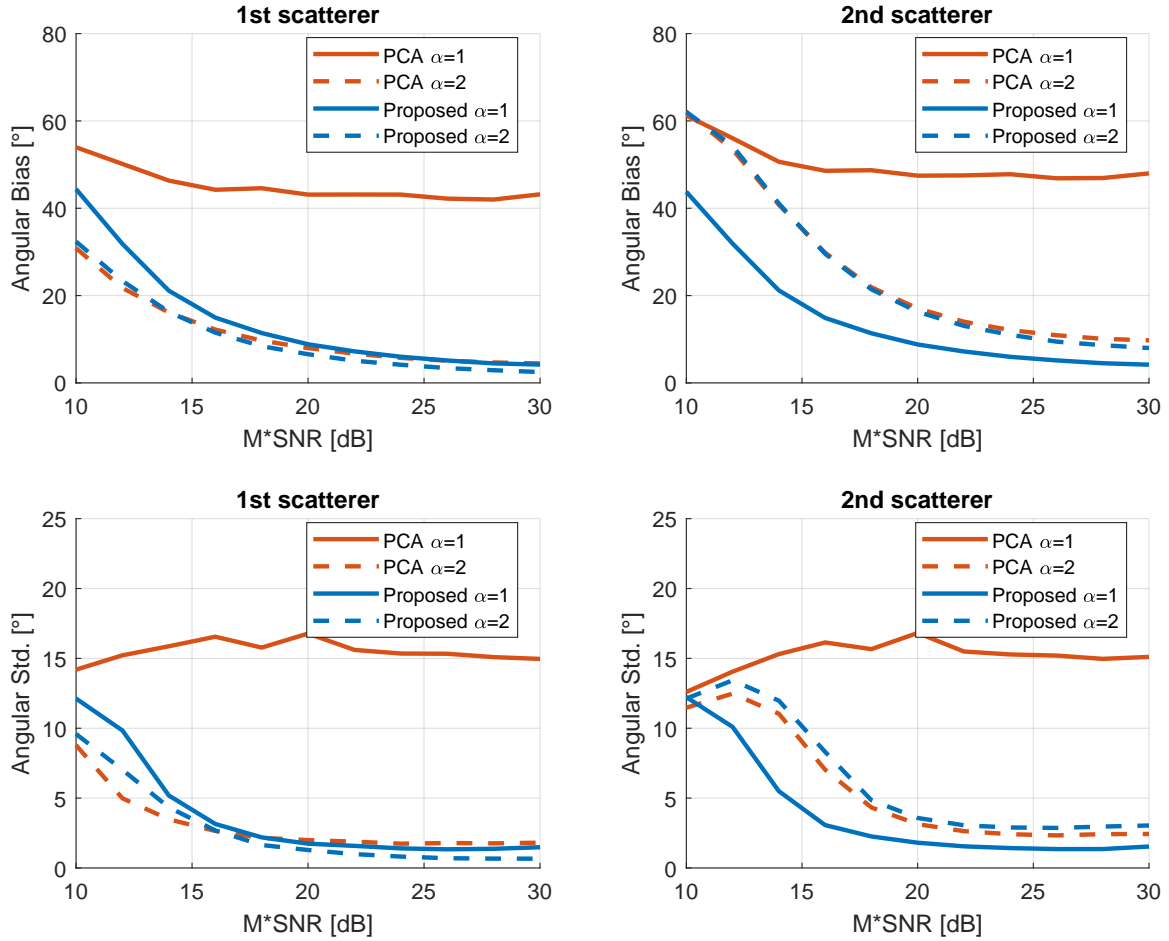


Fig. 7: The mean (upper row) and standard deviation (lower row) of the angular bias of the steering vector estimates with respect to  $M \times SNR$ . The solid curve corresponds to an amplitude ratio of 1, while the dashed curves correspond to a ratio of 2. The distance between the two scatterers was kept at one Rayleigh resolution. The result is consistent with Fig. 5 and Fig. 6. It is nearly impossible to separate two scatterers by PCA when the amplitude ratio is close to one, regardless of the SNR.

settings by the ensemble coherence (periodogram) of the first vector of  $\mathbf{Y}$  denoted by  $\mathbf{y}_1$  with the true steering vector of the first scatterer denoted by  $\mathbf{r}_1$ . As only the phase is of interest to us, the ensemble coherence (equation (23)) is computed based on the amplitude-dropped vectors of  $\mathbf{y}_1$  and  $\mathbf{r}_1$ , which are denoted by  $\bar{\mathbf{y}}_1$  and  $\bar{\mathbf{r}}_1$ .

$$\eta = \frac{1}{N} |\bar{\mathbf{y}}_1^H \bar{\mathbf{r}}_1| \quad (23)$$

The ensemble coherence  $\eta$  ranges from 0 to 1. A perfect reconstruction of the phase will result a coherence of one.

Fig. 10 (left) shows the ensemble coherence with respect to different polynomial orders applied to the simulated data used in Fig. 3 (with  $\alpha = 1.2$ ). The experiment shows that the ensemble coherence stays at almost one for a wide range of polynomial orders, which indicates a wide operable range. The coherence drops rapidly beyond 1.5, which aligns with our hypothesis in Section III-B that a higher polynomial order will introduce unwanted artificial scatterers with higher elevation. Our experiments show that a good starting point of the polynomial order is between 1.1 to 1.4. The operable range of  $\beta$  can be examined using the same strategy. Experiment in

Fig. 10 (right) shows that the ensemble coherence stays at nearly one beyond certain values of  $\beta$ . This demonstrates that the operable range of  $\beta$  is also very wide. In our experiments, we set  $\beta = 5$ .

Optionally, those parameters can also be adaptively estimated. The polynomial order  $d$  is mainly influenced by the elevation distance (angular difference) between the two scatterers because the polynomial order acts as a multiplication of the phase, whereas the parameter  $\beta$  mainly depends on the amplitude ratio of the two scatterers as the  $L_2$  norm in the Gaussian kernel is heavily governed by the signal magnitude. However, the elevation distance and amplitude ratio are both unknowns. One can start with an initial value of the kernel parameter, and refine it with the estimates of steering vectors and the amplitudes of the scatterers. In general, we found that a Gaussian kernel usually provides more stable performance than polynomial kernels in the experiments. The parameter  $\beta$  is also less sensitive to the change in the data.

#### B. Limitations of the proposed algorithm

First, the proposed algorithm assumes a fully-coherent signal model. We have shown in Section II-B that it does not

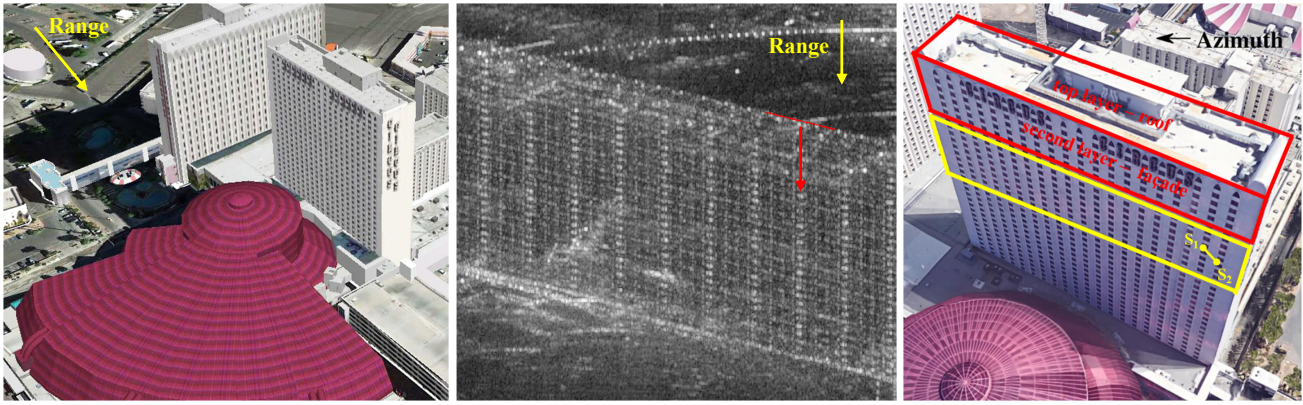


Fig. 8: Left: Google image of the test building; middle: the mean amplitude image of the test building; and right: the layover configuration of the test building (Fig. cited from [38]). The yellow arrows indicate the range direction. The red line on the amplitude image is a manually marked template of iso-height pixels. We assume that the pixels in the template have similar layover configuration, as well as similar scatterer statistics. It slides down one pixel at a time. We estimate a single reflectivity profile from the covariance matrix of the pixels in the template at a given position. In the experiment we also vary the length of the template to test the performance of the proposed algorithms under different number of looks.

work with the DS of an arbitrary coherence matrix. Additional constraints, such as coherence models, should be introduced in the proposed algorithm, in order to work with general DS models. However, our experiments do find the proposed algorithm works with a constant decorrelation matrix.

We also found that the sample selection may be a challenge in real data processing. This selection is different from many statistical tests mentioned in SqueeSAR or similar articles [44], [45], [46]. Therefore, we manually drew an iso-height template in the experiment. For a fully automatic processing, we shall incorporate available GIS building footprints, or resort to methods like NL-SAR [39]. However, these actions (especially NL-SAR) will increase the computational cost, which counteracts the motivation of the proposed algorithm.

Last, the proposed algorithm shows certain super-resolution capability. However, we must note that the knowledge of two scatterers was given in the experiments, meaning no model selection/detection was performed. Therefore, integrating a full detection step to the proposed algorithm is required for a complete assessment of its super-resolution power.

## VI. CONCLUSION

This article proposed a robust method to blindly perform layover separation in multibaseline InSAR data. Such blind separation requires no inversion of the SAR imaging model, hence reducing the computation cost logarithmically. The proposed algorithm outperforms the state of the art in various aspects. We showed that the state of the art could obtain a reasonable result only under good orthogonality conditions, i.e., large elevation and amplitude difference between the scatterers. The proposed method employs KPCA to artificially increase the dimension of the data, hence achieving superior performance. Simulation shows that the proposed algorithm is nearly optimum in the common range of SNR and number of looks. Experiments on real data show that the proposed method outperforms the state-of-the-art method by a factor of one to

three in terms of the height accuracy, depending on the used number of looks. Surprisingly, the proposed method is also capable of achieving a reasonable super-resolution capability, which is not shown in algorithms of the same kind.

This article shows a perspective on the data-driven approach of multibaseline InSAR algorithms. The long-term goal is to make good use of massive globally available SAR data. One immediate objective is to focus on an automatic and efficient sample selection strategies, such as incorporating freely available GIS building footprints. To further study data-driven approaches, we shall research subspace learning approaches in general, with application to TomoSAR. In addition, we can also investigate other emerging data-driven alternatives, such as deep learning to reduce the computational cost of TomoSAR parameter optimization.

## REFERENCES

- [1] G. Fornaro, D. Reale, and F. Serafino, "Four-dimensional sar imaging for height estimation and monitoring of single and double scatterers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 224–237, 1 2009, 00141.
- [2] X. Zhu and R. Bamler, "Very high resolution spaceborne sar tomography in urban environment," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 12, pp. 4296–4308, 2010, 00125.
- [3] X. X. Zhu and R. Bamler, "Let's do the time warp: Multicomponent nonlinear motion estimation in differential sar tomography," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 735–739, 2011, 00037.
- [4] D. Reale, G. Fornaro, and A. Pauciuillo, "Extension of 4-d sar imaging to the monitoring of thermally dilating scatterers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 12, pp. 5296–5306, 12 2013.
- [5] M. A. Siddique, U. Wegmüller, I. Hajnsek, and O. Frey, "Single-look sar tomography as an add-on to psi for improved deformation analysis in urban areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6119–6137, 10 2016.
- [6] A. Pauciuillo, D. Reale, A. De Maio, and G. Fornaro, "Detection of double scatterers in sar tomography," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 9, pp. 3567–3586, 9 2012.
- [7] P. Ma and H. Lin, "Robust detection of single and double persistent scatterers in urban built environments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 4, pp. 2124–2139, 4 2016.

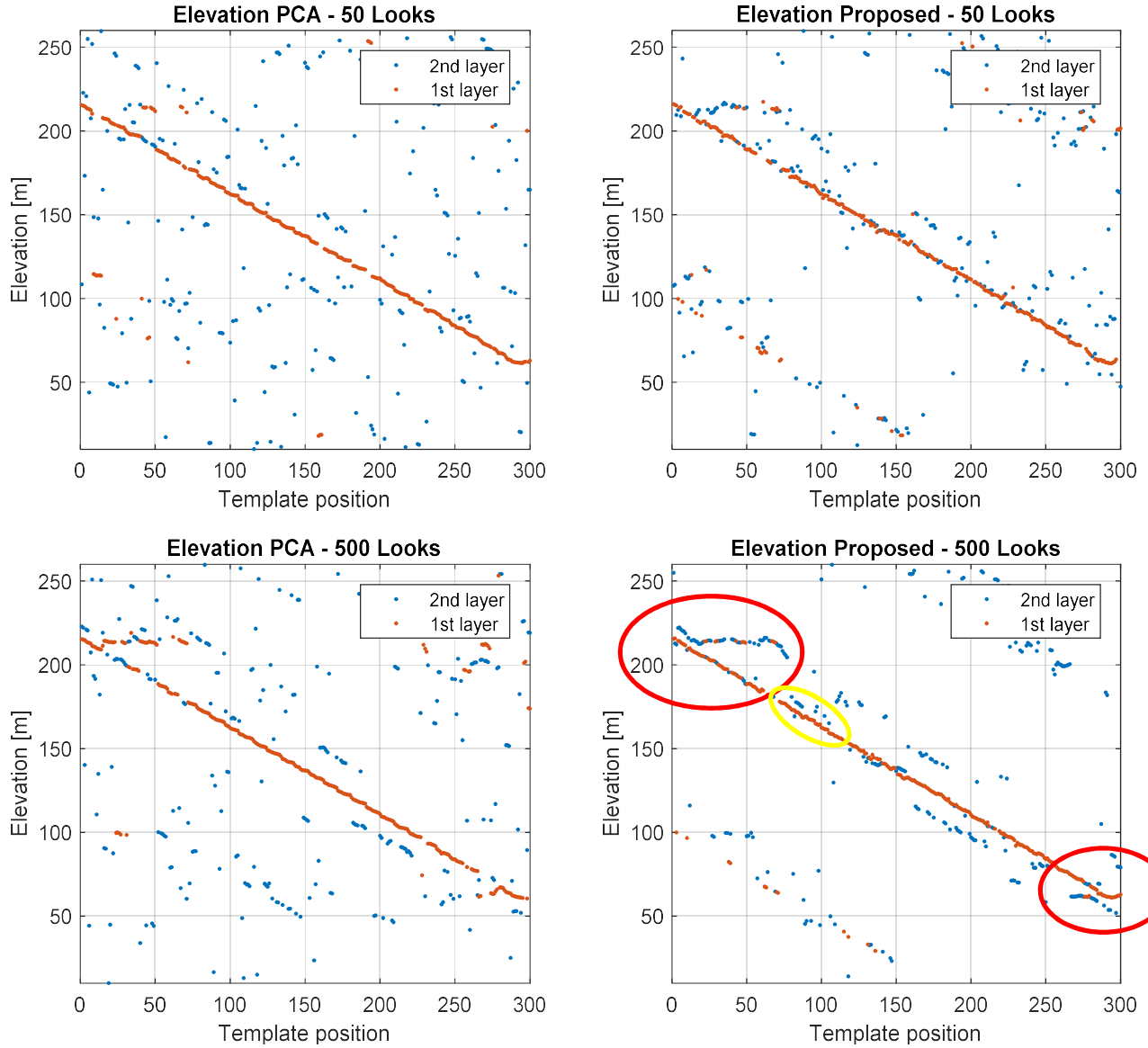


Fig. 9: Elevation retrieved from the first two principle directions estimated by PCA (left), and by proposed algorithm (right). The red dots represent the first (brightest) layer, and the blue is the second layer. The red and yellow ellipses mark two layover regions on the test building. They correspond to the drawing in Fig. 8 (right).

- [8] X. Zhu and R. Bamler, "Tomographic sar inversion by l1-norm regularization – the compressive sensing approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3839–3846, 2010, 00000.
- [9] X. X. Zhu and R. Bamler, "Superresolving sar tomography for multidimensional imaging of urban areas: Compressive sensing-based tomosar inversion," *Signal Processing Magazine, IEEE*, vol. 31, no. 4, pp. 51–58, 7 2014, 00002.
- [10] M. Eineder, C. Minet, P. Steigenberger, X. Cong, and T. Fritz, "Imaging geodesy - toward centimeter-level ranging accuracy with terrasars-x," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 2, pp. 661–671, 2 2011, 00060.
- [11] X. X. Zhu, S. Montazeri, C. Gisinger, R. Hanssen, and R. Bamler, "Geodetic sar tomography," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 18–35, 2015, 00002.
- [12] X. Zhu, *Very High Resolution Tomographic SAR Inversion for Urban Infrastructure Monitoring: A Sparse and Nonlinear Tour*, ser. Verlag der Bayerischen Akademie der Wissenschaften. Deutsche Geodätische Kommission, 2011, vol. 666, 00014.
- [13] G. Fornaro, F. Lombardini, A. Pauciuillo, D. Reale, and F. Viviani, "Tomographic processing of interferometric sar data: Developments, applications, and future research perspectives," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 41–50, 7 2014, 00003.
- [14] M. Bartusch, "Hrws high resolution wide swath: The next national x-band sar mission," TerraSAR-X/TanDEM-X Science Team Meeting, Oberpfaffenhofen, Germany, 2016.
- [15] G. Fornaro, S. Verde, D. Reale, and A. Pauciuillo, "Caesar: An approach based on covariance matrix decomposition to improve multibaseline-multitemporal interferometric sar processing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2050–2065, 4 2015, 00000.
- [16] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *Signal Processing, IEEE Transactions on*, vol. 45, no. 2, p. 434444, 1997, 02264.
- [17] J. F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, 10 1998, 02072.
- [18] J.-F. Cardoso, "Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem," *IEEE*, 1990, p. 26552658, 00179.



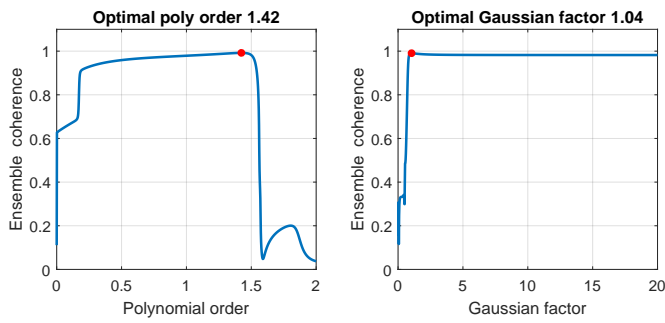


Fig. 10: The temporal coherence (complex correlation coefficient) between the first column of the dimension reduced data and the true steering vector of the brightest scatterer, with respect to different kernel parameter settings: order of polynomial kernel (left), and factor  $\alpha$  (in equation (18)) of the standard deviation of the Gaussian kernel (right). A coherence of 1 refers to a perfect reconstruction. In the simulation, two scatterers are separated by one Rayleigh resolution, and have an amplitude ratio of 1.2.

[19] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," vol. 140. IET, 1993, p. 362370, 02832.

[20] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," Springer, 1997, p. 583588, 01421.

[21] D. Martinez and A. Bray, "Nonlinear blind source separation using kernels," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 228–235, 1 2003, 00082.

[22] S. Harmeling, A. Ziehe, M. Kawanabe, and K.-R. Müller, "Kernel-based nonlinear blind source separation," *Neural Computation*, vol. 15, no. 5, p. 10891124, 2003, 00115.

[23] P. Honeine and C. Richard, "Preimage problem in kernel-based machine learning," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 77–88, 3 2011, 00058.

[24] J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural computation*, vol. 11, no. 1, p. 157192, 1999.

[25] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Transactions on signal processing*, vol. 50, no. 7, p. 15451553, 2002, 00393.

[26] T. Adali, C. Jutten, A. Yeredor, A. Cichocki, and E. Moreau, "Source separation and applications [from the guest editors]," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 16–17, 5 2014.

[27] S. R. Cloude and E. Pottier, "A review of target decomposition theorems in radar polarimetry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 2, pp. 498–518, 3 1996.

[28] N. Besic, G. Vasile, J. Chanussot, and S. Stankovic, "Polarimetric incoherent target decomposition by means of independent component analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 1236–1247, 3 2015, 00014.

[29] F. De Zan, "Optimizing sar interferometry for decorrelating scatterers," Ph.D. dissertation, Milan, Italy, 2008, 00007.

[30] Y. Wang and X. X. Zhu, "Robust Nonlinear Blind SAR Tomography in Urban Areas," in *EUSAR 2018; 12th European Conference on Synthetic Aperture Radar*, Jun. 2018, pp. 1–6, iSSN: null.

[31] M. A. Richards, J. A. Scheer, and W. A. Holm, Eds., *Principles of Modern Radar: Basic Principles*. Raleigh, NC: SciTech Publishing, 5 2010.

[32] I. T. Jolliffe, "Principal component analysis and factor analysis," in *Principal Component Analysis*, ser. Springer Series in Statistics. Springer, New York, NY, 1986, pp. 115–128.

[33] M. Even, "A study on spatio-temporal filtering in the spirit of squeeSar," *Proceedings of EUSAR 2016: 11th European Conference on Synthetic Aperture Radar*, 6 2016, pp. 1–4.

[34] J. F. Ayres, *Theory and Problems of Matrices : Including 340 Solved Problems, Completely Solved in Detail*, first edition edition ed. New York, NY: McGraw-Hill, 6 1967.

[35] T. Hofmann, B. Schölkopf, and A. J. Smola, "Kernel methods in machine learning," *The Annals of Statistics*, vol. 36, no. 3, pp. 1171–1220, 6 2008.

[36] Q. Wang, "Kernel principal component analysis and its applications in face recognition and active shape models," *arXiv:1207.3538 [cs]*, 7 2012, arXiv: 1207.3538.

[37] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*, 1st ed. Upper Saddle River, N.J: Prentice Hall, 2 1997.

[38] X. X. Zhu, N. Ge, and M. Shahzad, "Joint sparsity in sar tomography for urban mapping," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 8, pp. 1498–1509, 12 2015, 00000.

[39] C.-A. Deledalle, L. Denis, F. Tupin, A. Reigber, and M. Jager, "Nl-sar: A unified nonlocal framework for resolution-preserving (pol)(in)sar denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2021–2038, 4 2015, 00015.

[40] Y. Wang and X. X. Zhu, "Robust Estimators for Multipass SAR Interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 2, pp. 968–980, Feb. 2016, 00000.

[41] S. Visuri, H. Oia, and V. Koivunen, "Nonparametric statistics for subspace based frequency estimation," in *2000 10th European Signal Processing Conference*, Sep. 2000, pp. 1261–1264.

[42] A. Ferretti, C. Prati, and F. Rocca, "Permanent scatterers in SAR interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 1, pp. 8–20, Jan. 2001, 01728.

[43] X. X. Zhu and R. Bamler, "Let's Do the Time Warp: Multicomponent Nonlinear Motion Estimation in Differential SAR Tomography," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 735–739, 2011, 00037.

[44] A. Ferretti, A. Fumagalli, F. Novali, C. Prati, F. Rocca, and A. Rucci, "A new algorithm for processing interferometric data-stacks: Squeesar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 9, pp. 3460–3470, 9 2011, 00204.

[45] A. Parizzi and R. Bricic, "Adaptive insar stack multilooking exploiting amplitude statistics: A comparison between different techniques and practical results," *Geoscience and Remote Sensing Letters, IEEE*, vol. 8, no. 3, pp. 441–445, 5 2011, 00049.

[46] Y. Wang, X. Zhu, and R. Bamler, "Retrieval of phase history parameters from distributed scatterers in urban areas using very high resolution sar data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 73, pp. 89–99, 9 2012, 00000.



**Yuanyuan Wang** received the B.Eng. degree (Hons.) in electrical engineering from The Hong Kong Polytechnic University, Hong Kong, in 2008, and the M.Sc. and Dr. Ing. degree from the Technical University of Munich (TUM), Munich, Germany, in 2010 and 2015, respectively. In June and July 2014, he was a Guest Scientist with the Institute of Visual Computing, ETH Zurich, Zurich, Switzerland. He is currently with the Department of EO Data Science, Remote Sensing Technology Institute, German Aerospace Center, Weling, Germany, where he leads the working group Big SAR Data. His research interests include optimal and robust parameters estimation in multibaseline InSAR techniques, multisensor fusion algorithms of synthetic aperture radar (SAR) and optical data, nonlinear optimization with complex numbers, machine learning in SAR, and high-performance computing for big data. Dr. Wang was one of the best reviewers of the IEEE Transactions on Geoscience and Remote Sensing in 2016. He is a Member of the IEEE.





**Xiao Xiang Zhu** (S'10–M'12–SM'14) received the Master (M.Sc.) degree, her doctor of engineering (Dr.-Ing.) degree and her Habilitation in the field of signal processing from Technical University of Munich (TUM), Munich, Germany, in 2008, 2011 and 2013, respectively.

She is currently the Professor for Signal Processing in Earth Observation ([www.sipeo.bgu.tum.de](http://www.sipeo.bgu.tum.de)) at Technical University of Munich (TUM) and the Head of the Department “EO Data Science” at the Remote Sensing Technology Institute, German Aerospace Center (DLR). Since 2019, Zhu is a co-coordinator of the Munich Data Science Research School ([www.mu-ds.de](http://www.mu-ds.de)). Since 2019 She also heads the Helmholtz Artificial Intelligence Cooperation Unit (HAICU) – Research Field “Aeronautics, Space and Transport”. Since May 2020, she is the director of the international future AI lab “AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond”, Munich, Germany. Prof. Zhu was a guest scientist or visiting professor at the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China, the University of Tokyo, Tokyo, Japan and University of California, Los Angeles, United States in 2009, 2014, 2015 and 2016, respectively. Her main research interests are remote sensing and Earth observation, signal processing, machine learning and data science, with a special application focus on global urban mapping.

Dr. Zhu is a member of young academy (Junge Akademie/Junges Kolleg) at the Berlin-Brandenburg Academy of Sciences and Humanities and the German National Academy of Sciences Leopoldina and the Bavarian Academy of Sciences and Humanities. She is an associate Editor of IEEE Transactions on Geoscience and Remote Sensing.